

Molecular Evolution of Hepatitis B Virus over 25 Years

Carla Osiowy,^{1*} Elizabeth Giles,¹ Yasuhito Tanaka,² Masashi Mizokami,² and Gerald Y. Minuk³

*National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada¹;
Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of
Medical Sciences, Kawasumi, Mizuho, Nagoya, Japan²; and Section of Hepatology, Department of
Medicine, University of Manitoba, Winnipeg, Manitoba, Canada³*

Received 15 May 2006/Accepted 3 August 2006

Determining the longitudinal molecular evolution of hepatitis B virus (HBV) is difficult due to HBV's genomic complexity and the need to study paired samples collected over long periods of time. In this study, serial samples were collected from eight hepatitis B virus e antigen-negative asymptomatic carriers of HBV genotype B in 1979 and 2004, thus providing a 25-year period to document the long-term molecular evolution of HBV. The rate, nature, and distribution of mutations that emerged over 25 years were determined by phylogenetic and linear regression analysis of full-length HBV genome sequences. Nucleotide hypervariability was observed within the polymerase and pre-S/S overlap region and within the core gene. The calculated mean number of nucleotide substitutions/site/year (7.9×10^{-5}) was slightly higher than published estimates (1.5×10^{-5} to 5×10^{-5}). Nucleotide changes in the quasispecies population did not significantly alter the molecular evolutionary rate, based on linear regression analysis of evolutionary distances among serial clone pre-S region sequences. Therefore, the directly amplified or dominant sequence was sufficient to estimate the putative molecular evolutionary rate for these long-term serial samples. On average, the ratio of synonymous (d_s) to nonsynonymous (d_N) substitutions was highest for the polymerase-coding region and lowest for the core-coding region. The low d_s/d_N ratios observed within the core suggest that selection occurs within this gene region, possibly as an immune evasion strategy. The results of this study suggest that HBV sequence divergence may occur more rapidly than previously estimated, in a host immune phase-dependent manner.

Determination of the molecular evolution of hepatitis B virus (HBV) involves an understanding of the accumulated sequence changes to the viral genome and the observed mutation rate over a long period. Determining the rate of sequence change is difficult due to the complex organization of the HBV genome, which involves multiple coding and regulatory functions within overlapping open reading frames (16). Two-thirds of the viral genome codes for multiple proteins, and thus a synonymous change in one open reading frame results in a nonsynonymous change in the overlapping open reading frame. In this way, it is believed that HBV genome evolution is “constrained” in order to maintain essential protein functions (22).

Since HBV replication involves an error-prone reverse transcription step, the rate of nucleotide change during replication is higher than that found for other DNA viruses and is more similar to the rate observed for the slower-evolving RNA viruses (21, 24). The rate of HBV evolution in hepatitis B virus e antigen (HBeAg)-positive individuals has been estimated to be 1.5×10^{-5} to 5×10^{-5} nucleotide substitutions per site per year (1, 13, 24, 29). However, the mutation rate or total accumulated number of mutations appears to be higher in HBeAg-negative patients (11, 33), suggesting that the host immune response plays an important role in HBV evolution.

There are very few reports of long-term longitudinal studies involving the full-length HBV genome, likely due to the diffi-

culty in collecting paired samples from individual chronic HBV patients over a long period (11). The present study analyzed HBV genome sequence changes occurring over time (25 years) in eight HBeAg-negative patients. The molecular evolutionary rate and distribution of mutations occurring throughout the entire genome and in the pre-S gene region of serial sample quasispecies were also investigated.

MATERIALS AND METHODS

Patients. The eight hepatitis B virus surface antigen (HBsAg)-positive subjects described in this study were identified as a result of their participation in a large seroepidemiologic survey of HBV infection within their community conducted in 1979 (20). Twenty-one additional subjects were also found to be HBsAg positive during that survey, but these individuals had either died or left the settlement or were out “on the land” hunting when investigators returned in 2004 to repeat clinical evaluations and obtain follow-up serum samples. None of the eight participating subjects had received HBV treatment or HBV immunoprophylaxis or had been treated with immunosuppressant drugs for other medical disorders during the intervening 25 years. Informed consent was obtained from each subject on both occasions. The two study protocols (1979 and 2004) were approved by the University of Manitoba Conjoint Ethics Committee for Human Experimentation.

DNA extraction. DNA was extracted from 150 μ l of serum by the proteinase K-sodium dodecyl sulfate lysis and phenol-chloroform extraction methods, as described previously (25), and was resuspended (final volume, 30 μ l) in sterile, nuclease-free water. The extracted DNA was stored at -20°C .

Viral load determination. HBV DNA quantitation was performed by real-time PCR analysis using a RealArt HBV PCR kit (Artus Biotech, QIAGEN, Mississauga, Ontario, Canada) with an ABI Prism 7500 sequence detection system. Briefly, 2.5 μ l of DNA extract was added to 7.5 μ l sterile water and 15 μ l kit master mix consisting of buffers, enzyme, primers, and probe for the specific amplification of a 134-bp region of the HBV genome. One microliter of kit internal control was also added per reaction to identify possible PCR inhibition. The DNA quantity (international units [IU]/ml) was determined by comparison to external quantitation standards (range, 10 IU to 1×10^5 IU). Real-time PCR

* Corresponding author. Mailing address: Public Health Agency of Canada, National Microbiology Laboratory, Canadian Science Centre for Human and Animal Health, 1015 Arlington Street, Winnipeg, Manitoba R3E 3P6, Canada. Phone: (204) 789-6061. Fax: (204) 789-2082. E-mail: carla_osiowy@phac-aspc.gc.ca.

TABLE 1. Sequencing primers used for this study

| Set | Primer name | Sequence (5'-3') | Annealing temp (°C) | Expected size (bp) |
|-----|-------------|-------------------------------|---------------------|--------------------|
| 1 | FLG1 | CAC CTC TGC CTA ATC ATC | 50 | 3,210 |
| | FLG2 | GTT GCA TGG TGC TGG TC | | |
| 2 | Pol F1 | CYT TYG GAG TGT GGA TTC GC | 55 | 1,566 |
| | Pol R1 | TGG GAT GGG AAT ACA RGT GC | | |
| 3 | Pol F2 | CGT TTG TCC TCT AMT TCC AGG | 55 | 961 |
| | Pol R2 | ACG TAR ACA AAG GAC GTC CC | | |
| 4 | X F2 | TTG CTC GCA GCM GGT CTG GAG C | 55 | 525 |
| | FLG2 | See above | | |
| 5 | FLG1 | See above | 55 | 651 |
| | Core 2R | YCC CAC CTT ATG WTG CCA AGG | | |

cycling parameters and result interpretation were carried out according to the manufacturer's protocol.

PCR amplification. Full-length genome sequencing of HBV DNA was performed by nested PCR with a full-length amplicon obtained using the primers and thermocycling conditions described by Günther et al. (10). Thereafter, several nested PCR steps were performed in order to increase the sensitivity of detection and produce a sufficient amount of amplicon for sequence analysis. The sequences and annealing temperatures of the nested PCR primer sets used are shown in Table 1. PCR was performed using an ultra-high-fidelity polymerase (AccuPrime *Pfx* DNA polymerase; Invitrogen Life Technologies, Burlington, Ontario, Canada) to ensure low to nil error rates during amplification (5). Reaction tubes for PCR contained 5 µl DNA extract or 2 µl of the first-stage PCR product, AccuPrime *Pfx* reaction mix (Invitrogen Life Technologies), a 0.5 µM concentration of each primer, and 1 U AccuPrime *Pfx* DNA polymerase. Thermal cycling parameters for each set of primers were those suggested by the manufacturer (Invitrogen Life Technologies) for three-step cycling using the annealing temperatures listed in Table 1.

Full-genome sequence analysis. Nested PCR products were gel purified prior to cycle sequencing with an ABI Prism 3100 genetic analyzer (Applied Biosystems, Foster City, California), using BigDye v3.1 Terminator chemistry. All sequences were assembled using SeqMan II software (DNASTar Inc., Madison, Wisconsin). Full-length genome sequence alignments were performed using ClustalX v1.8 (32). Sequence identity and divergence were calculated based on the number of nucleotide changes per total number of nucleotides analyzed (3,215 bp). The number of nucleotide substitutions per site per year, based on direct comparison between sample pair sequences, was calculated using the equation of Gojobori and Yokoyama (8, 24). Phylogenetic tree analysis was performed using the Tamura-Nei model of evolutionary distance, and the topology was evaluated by bootstrap analysis (1,000 replicates) using the neighbor-joining method. Linear regression analyses based on evolutionary distances obtained with the Tamura-Nei model for all positions were performed to calculate the final mean molecular evolutionary rate for full-length genome sequences (30). Phylogenetic analysis and evolutionary distance calculations were performed using MEGA v2.1 software (15). The ratio of synonymous (d_s) to non-synonymous (d_n) substitutions for protein-coding regions among matched sam-

ples was calculated by the method of Nei and Gojobori (23), using SNAP software (14; www.hiv.lanl.gov).

Quasispecies analysis. Viral quasispecies were investigated by clonal analysis of the pre-S gene regions from four sample pairs. A 479-bp amplicon from the pre-S1/pre-S2 gene region was obtained, using primers P1 and P2 as described previously (17). Amplicons were gel purified and cloned into a pCR2.1-TOPO plasmid vector (Invitrogen Life Technologies) according to the manufacturer's instructions. Ligated products were transformed into *Escherichia coli* TOP10F cells (Invitrogen Life Technologies), at least 10 individual colonies were picked, and the plasmid DNA inserts were sequenced.

Mean genetic distances of all synonymous and nonsynonymous positions were calculated using the Pamilo-Bianchi-Li (P-B-L) model (MEGA v2.1). Linear regression analyses based on Tamura-Nei model evolutionary distances were performed to calculate a mean evolutionary rate of quasispecies sequences over the 25-year period.

Nucleotide sequence accession numbers. The full-length genome sequences obtained from the patients at each time point were submitted to the National Center for Biotechnology Information GenBank database under accession numbers DQ463787 to DQ463802.

RESULTS

Patient clinical and virological data. As shown in Table 2, the mean (\pm standard deviation) age of the eight HBsAg-positive subjects in 2004 was 69.8 ± 12.6 years, and 7/8 (88%) subjects were male. All subjects were HBeAg negative and anti-HBe positive on both occasions (1979 and 2004). Seven subjects were positive for antibody to hepatitis A virus (anti-HAV) and negative for anti-HCV in both 1979 and 2004. The remaining individual was anti-HAV negative in 1979 but positive in 2004. This individual was also anti-HCV negative on both occasions. Liver biochemistry (serum alanine aminotrans-

TABLE 2. Characteristics of the patient population^a

| Patient no. | Age (yr) in 1979/2004 | Gender | HBV DNA (IU/ml [10^3]) | | Alanine aminotransferase level | | Alkaline phosphatase level | | Total bilirubin | | Albumin level | |
|-------------|-----------------------|--------|----------------------------|------|--------------------------------|------|----------------------------|------|-----------------|------|---------------|------|
| | | | 1979 | 2004 | 1979 | 2004 | 1979 | 2004 | 1979 | 2004 | 1979 | 2004 |
| 524-3 | 38/63 | Male | 0.3 | 0.9 | 31 | 30 | 119 | 13 | 0.2 | 7 | 4.7 | 44 |
| 462-4 | 54/79 | Female | 2.3 | 0.4 | 4 | 30 | 96 | 61 | 0.2 | 9 | 4.5 | 43 |
| 234-6 | 30/55 | Male | 8.8 | 7.1 | 24 | 36 | 86 | 65 | 0.1 | 6 | 4.4 | 43 |
| 928-9 | 59/84 | Male | 1.0 | 3.4 | 21 | 17 | 61 | 64 | 0.2 | 7 | 4.4 | 43 |
| 739-11 | 24/49 | Male | 2.8 | 2.0 | 18 | 52 | 93 | 132 | NA | 6 | NA | 43 |
| 650-15 | 41/66 | Male | 0.2 | 3.8 | 15 | 30 | 78 | 75 | 0.4 | 11 | 4.8 | 37 |
| 539-16 | 59/84 | Male | 2.4 | 1.1 | 17 | 23 | 95 | 98 | 0.2 | 7 | 4.5 | 40 |
| 621-14 | 53/78 | Male | 1.3 | 0.1 | 121 | 101 | 121 | 101 | 0.7 | 10 | 4.7 | 42 |

^a All patients were positive for HBsAg in both 1979 and 2004. Normal laboratory values were as follows in 1979 and 2004, respectively: alanine aminotransferase, 2 to 35 MU/ml and 21 to 72 IU/liter; alkaline phosphatase, 30 to 130 MU/ml and 30 to 136 IU/liter; total bilirubin, 0.2 to 1 mg/dl and 3 to 22 µmol/liter; and albumin, 3.5 to 5 g/dl and 35 to 50 g/liter. NA, not available.

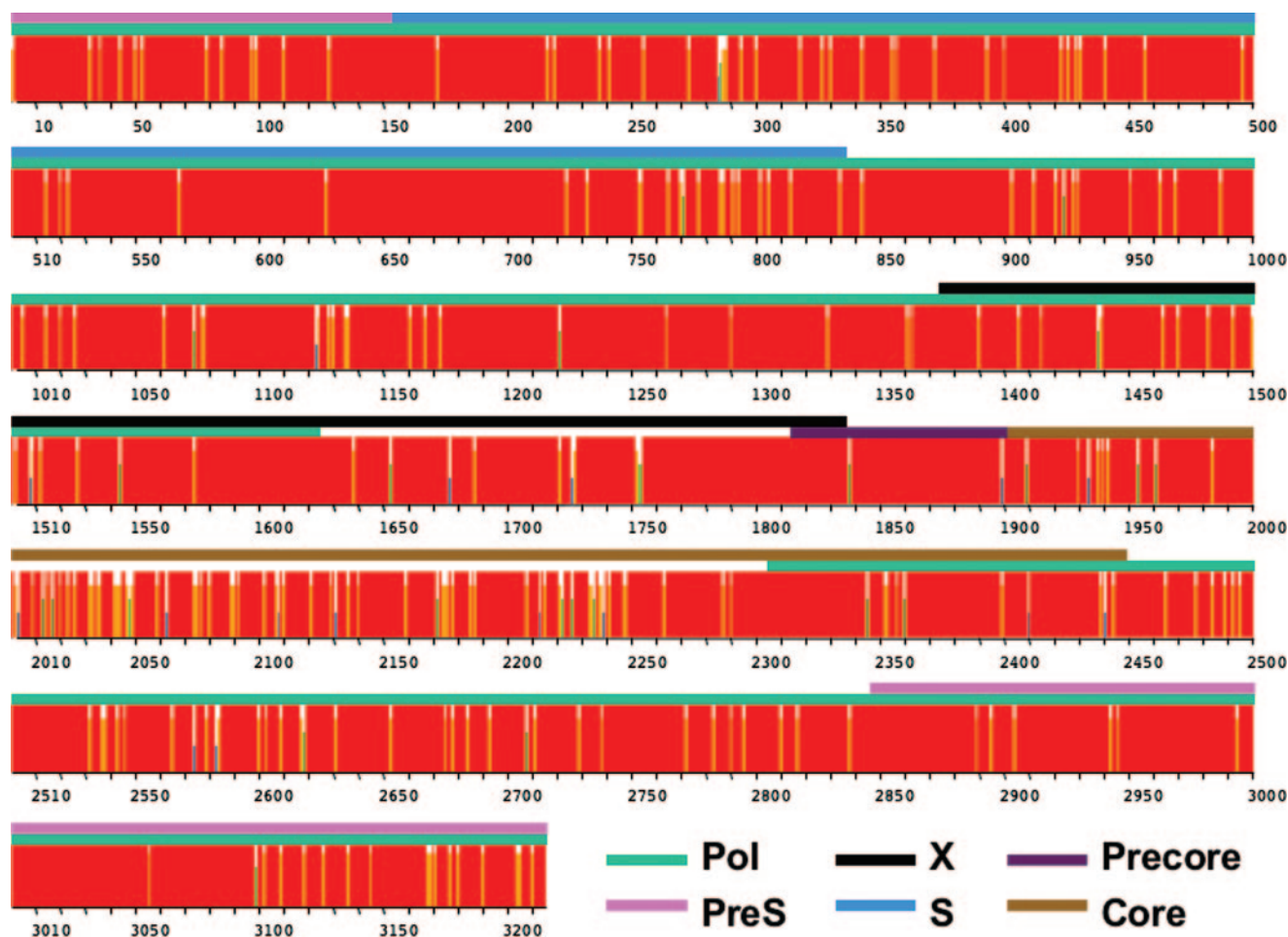


FIG. 1. Alignment histogram of 16 full-length HBV genome sequences. The different colored bars above the histogram denote the coding regions for the indicated proteins. The ruler below the histogram denotes the numbers of nucleotides along the length of the genome. Nucleotide substitutions compared to the consensus are shown as colored slashes within the red consensus bar, with increasing diversity shown as follows: orange < green < blue.

ferase, alkaline phosphatase, total bilirubin, and albumin levels) tests were normal for all individuals on each occasion. HBV DNA viral loads were variable for most patients at the two time points measured, ranging over approximately 1 log. Overall, the DNA levels remained very low at both time points for all patients ($<10^4$ IU/ml).

Sequencing. All sera collected in 2004 were extracted, amplified, and sequenced prior to analysis of the 1979 sample set to avoid the possibility of contamination between matched samples. Following sequencing and assembly of each genome sequence, a 3,215-bp genome was obtained for each sample. Each genome sequence was genotyped using the NCBI genotyping tool (26), and all were determined to be genotype B. One sample pair (from patient 462-4) had a premature stop codon within HBcAg (182 of 184 codons) and lacked a start codon for the pre-S2 gene for the viral genomes from both time points. The precore stop mutation at nucleotide 1896 of the HBcAg gene was observed in all genome sequences; however, none of the genomes contained mutations at nucleotides 1762 and 1764 within the core promoter region.

Phylogenetic analysis. The 16 full-length HBV genome sequences were aligned, and a histogram was prepared to visually demonstrate regions of hypervariability and relative conservation along the length of the genome (Fig. 1). Nucleotide changes occurred throughout the entire genome and in each coding region. Regions of hypervariability were observed within the core gene, the 3' end region of the S gene, and other regions of overlap between the polymerase and pre-S/S genes. Conversely, the X gene and the overlap region encompassing the S major hydrophilic region and reverse transcriptase domains B and C within the polymerase gene were observed to have fewer nucleotide substitutions among the 16 sequences.

Phylogenetic analysis showed that most matched samples formed unique clusters within the tree, while two sample pairs did not cluster significantly, suggesting that the matched samples from these pairs diverged independently (Fig. 2). In order to avoid a false overestimation of the mean molecular evolutionary rate, sequences from these two sample pairs were not included in rate calculations, except for quasispecies analysis (for patient 234-6 only). The range of sequence divergence

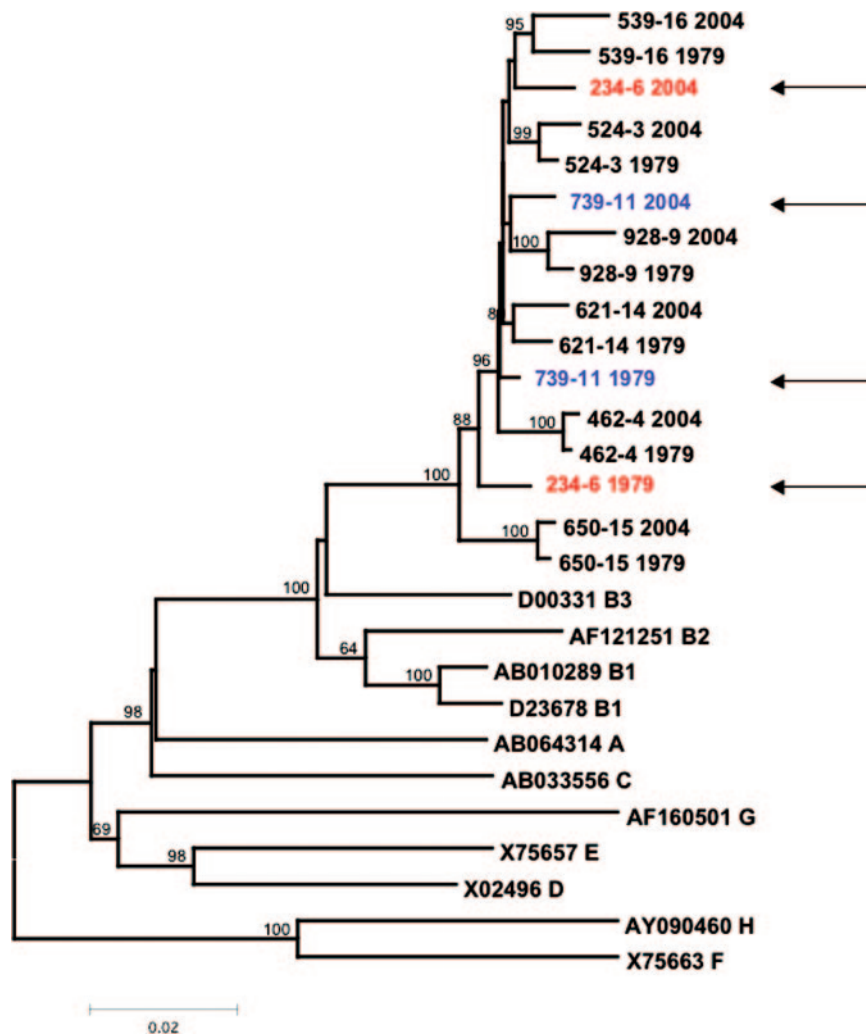


FIG. 2. Phylogenetic analysis of 16 full-length HBV genome sequences. Arrows indicate the two sample pairs showing independent evolution (234-6 [red] and 739-11 [blue]). Sequence alignment was performed using ClustalX v.1.8. Evolutionary distances were calculated using the Tamura-Nei model, and the phylogenetic tree topology was evaluated by bootstrap analysis (1,000 replicates) using the neighbor-joining method (confidence values of 50% or greater are shown). The ruler shows the branch length for a pairwise distance equal to 0.02.

among the eight matched pairs was 0.3% (patient 462-4) to 1.9% (patient 234-6).

Molecular evolutionary rate analysis. Values for the number of nucleotide substitutions per site per year, based on direct comparison between the six sample pair sequences that uniquely clustered, ranged from 6.23×10^{-5} (patient 462-4) to 3.59×10^{-4} (patient 539-16), with a mean value of 1.9×10^{-4} . To more accurately determine the molecular evolutionary rate based on a putative ancestral sequence, linear regression analysis was performed (Fig. 3). Using this method of analysis, the mean evolutionary rate among six sample pairs was determined to be 7.9×10^{-5} nucleotide substitutions/site/year.

Quasispecies analysis. The contribution of viral quasispecies to the molecular evolutionary rate of sample pairs over the 25-year period was investigated by clonal analysis of four sample pairs. Sample pairs were chosen based on their divergence over the study period, with two pairs demonstrating the most sequence divergence over time (234-6 and 539-16 [1.9% and 1.8% divergence, respectively]) and two

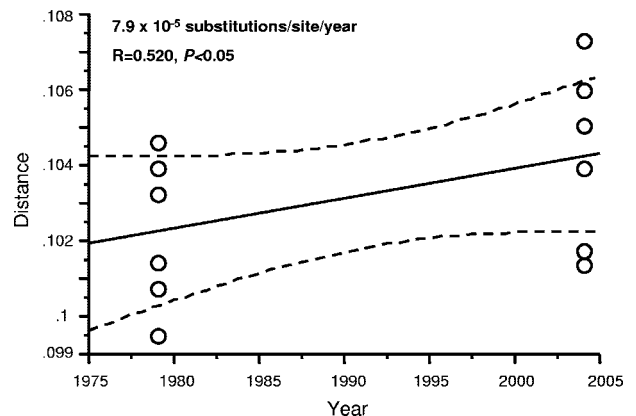


FIG. 3. Molecular evolutionary rate estimation based on long-term serial HBV samples from HBeAg-negative asymptomatic carriers of HBV genotype B. The results of regression analysis of evolutionary distances among the six sample pair sequences demonstrating significant phylogenetic clustering are shown. The mean evolutionary rate (solid regression line) from the Tamura-Nei model is indicated above the graph. The 95% confidence intervals of the regression line are indicated by dashed lines.

TABLE 3. Quasispecies genetic distance analysis of four sample pairs over 25 years

| Sample pair | 1979 values | | | | 2004 values | | | |
|-------------|-------------|---------|------------------|-----------|-------------|---------|---------|-----------|
| | d_N^a | d_S^b | All ^c | d_N/d_S | d_N | d_S | All | d_N/d_S |
| 234-6 | 0.00755 | 0.00619 | 0.00648 | 1.21971 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| 462-4 | 0.00099 | 0.00204 | 0.00199 | 0.48529 | 0.00084 | 0.00000 | 0.00056 | 0.00000 |
| 539-16 | 0.00149 | 0.00281 | 0.00381 | 0.53025 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| 650-16 | 0.00337 | 0.00504 | 0.00363 | 0.66865 | 0.00155 | 0.01369 | 0.00448 | 0.11322 |

^a Mean genetic distance of all quasispecies nonsynonymous positions calculated using the P-B-L model.

^b Mean genetic distance of all quasispecies synonymous positions calculated using the P-B-L model.

^c Mean genetic distance of all positions for all quasispecies calculated using the Tamurai-Nei model.

pairs demonstrating the least sequence divergence over time (462-4 and 650-15 [0.3% and 0.4% divergence, respectively]). The pre-S region was selected for quasispecies analysis because this region is recognized for its hypervariability within the HBV genome (13).

The mean overall genetic distances (including synonymous and nonsynonymous sites) between quasispecies sequences from 1979 and 2004 were determined (Table 3). A significant reduction in the overall genetic distances in 2004 compared to those in 1979 was observed (except for samples from patient 650-16), suggesting that selective pressure occurred within the pre-S region. Following nonsynonymous versus synonymous substitution analysis of quasispecies, only sample pair 234-6 quasispecies showed greater nonsynonymous than synonymous changes over time (ratio, 1.247), indicating positive selection. The other three sample pairs showed no positive selection over the 25-year period (ratios of nonsynonymous to synonymous substitutions, 0.438 [462-4], 0.492 [539-16], and 0.232 [650-16]). Sample pair 234-6 quasispecies sequences also showed a shift within the pre-S2 region, with a three-codon deletion in 6 of 11 clones from 2004 that was not observed in any clones from 1979 ($n = 11$).

The mean molecular evolutionary rate based on pre-S quasispecies sequences from the four sample pairs (8.3×10^{-5} /site/year) (Fig. 4A) was compared to the calculated rate based on the pre-S region from each dominant sequence of the six full-

length sample pairs as well as the samples from patient 234-6 (7.2×10^{-5} /site/year) (Fig. 4B). No significant difference in mean rate was observed ($P > 0.05$), indicating that quasispecies variation does not contribute considerably to the overall mean evolutionary rate during long-term follow-up over 25 years.

Synonymous versus nonsynonymous substitutions of full-length genome coding regions. The ratio (d_S/d_N) of synonymous (d_S) to nonsynonymous (d_N) substitutions was calculated for the HBsAg, polymerase, core, and X coding regions for the six sample pairs showing unique phylogenetic clustering (Table 4). This ratio determines the extent of natural selection, such that a ratio of <1 indicates positive selection within the gene. Ratios of <1 were observed for several coding regions from several of the patients, but the majority of coding regions had ratios of >1 , indicative of sequence stability or negative selection over time. In general, the lowest ratios were observed for the core coding region. Conversely, the polymerase coding region demonstrated the highest d_S/d_N ratios among all patients.

DISCUSSION

The current study investigates the molecular evolution of HBV in individuals over a 25-year period. To our knowledge, this study is one of the first to analyze longitudinal HBV

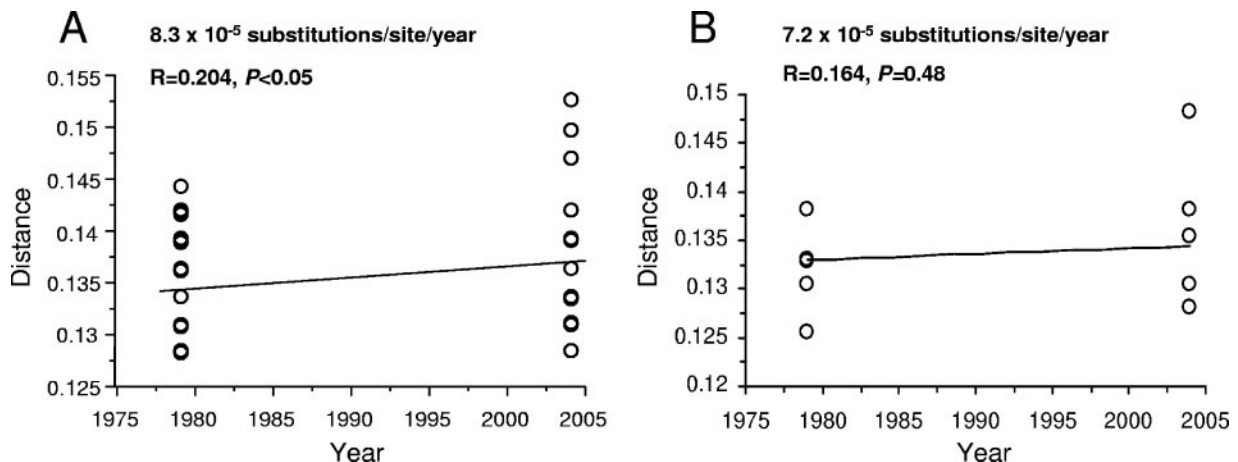


FIG. 4. Comparison of evolutionary rate calculations based on quasispecies versus dominant strain sequences. Evolutionary rates were calculated by regression analyses (Tamura-Nei model) using either (A) quasispecies (cloned) pre-S sequences ($n = 114$) from four sample pairs or (B) dominant strain (directly amplified) pre-S sequences ($n = 14$) from seven sample pairs. The differences between the evolutionary rates in panels A and B are not significant ($P > 0.05$).

TABLE 4. Ratio of synonymous to nonsynonymous substitutions for coding regions among full-length genomes

| Sample pair | Coding region ^d | | | | | | | | | | | |
|-------------|----------------------------|------------------------|---------------------------|-----------|-----------|--------------|-----------|-----------|--------------|-----------|-----------|--------------|
| | HBsAg | | | Pol | | | HBcAg | | | X | | |
| | <i>Sd</i> ^a | <i>Sn</i> ^b | <i>Sd/Sn</i> ^c | <i>Sd</i> | <i>Sn</i> | <i>Sd/Sn</i> | <i>Sd</i> | <i>Sn</i> | <i>Sd/Sn</i> | <i>Sd</i> | <i>Sn</i> | <i>Sd/Sn</i> |
| 621-14 | 2.00 | 1.00 | 6.56 | 6.00 | 17.00 | 1.12 | 4.00 | 11.00 | 1.13 | 3.50 | 3.50 | 3.01 |
| 524-3 | 1.00 | 2.00 | 1.64 | 10.00 | 6.00 | 5.31 | 2.00 | 6.00 | 1.04 | 2.00 | 4.00 | 1.45 |
| 650-15 | 0.00 | 8.00 | NA | 8.00 | 4.00 | 6.37 | 0.00 | 1.00 | NA | 1.00 | 0.00 | NA |
| 539-16 | 5.00 | 10.00 | 1.65 | 23.00 | 13.00 | 5.70 | 6.00 | 10.00 | 1.91 | 1.50 | 5.50 | 0.80 |
| 462-4 | 1.00 | 1.00 | 3.29 | 1.00 | 3.00 | 1.05 | 1.00 | 3.00 | 1.01 | 0.00 | 2.00 | NA |
| 928-9 | 2.00 | 8.00 | 0.81 | 12.50 | 9.50 | 4.20 | 4.00 | 9.00 | 1.39 | 3.00 | 2.00 | 4.42 |

^a Number of observed synonymous substitutions.^b Number of observed nonsynonymous substitutions.^c Ratio of synonymous to nonsynonymous substitutions, calculated using the Jukes-Cantor correction for multiple hits of *ps* and *pn* (23).^d NA, not available (*Sd/Sn* was not calculated if either *Sd* or *Sn* equaled 0).

molecular evolution by using paired samples collected over a long period of time from the same patients. Due to the difficulty in sampling individuals over such extended periods of time, most studies have compared genome sequences to a “virtual” baseline sequence comprised of a composite or consensus sequence from current viral quasiespecies or the viral sequence assumed to be transmitted (i.e., the mother’s current HBV sequence for mother-to-child transmission) (2, 11, 24). In the present study, the molecular evolution of HBV strains assumed to have undergone very little selection pressure following an intense period of host immune responses during seroconversion could be studied, as only treatment-naïve, asymptomatic HBeAg-negative carriers were investigated. Previous studies have focused on the evolution of HBV in HBeAg-positive carriers, who have a considerably different immune profile. Therefore, changes to the viral genome observed in this study over the 25-year period are more likely a true representation of mutational changes due to viral polymerase errors and the selective outgrowth of fit variants than to selective pressure due to host immune responses. Hence, we believe that our data give an accurate representation of the distribution of mutations occurring within the viral genome over time and are more reliable for the accurate determination of the molecular evolution and molecular clock of HBV.

An alignment of all genomic sequences demonstrated the presence of nucleotide substitutions throughout the entire genome. Their distribution was not entirely even along the length of the genome, with regions of apparent clustering or absence of substitutions, suggesting that certain proteins may have greater importance as immune targets. This contrasts with the findings of Hannoun et al. (11), who found the entire HBV genome to be extremely stable, with mutations distributed fairly evenly in all coding regions, particularly in HBeAg-positive patients. Some of the observed substitutions may also be intragenotypic variations within genotype B viruses and not necessarily associated with natural selection. Sequence similarity among all matched pairs was >98%, which is well within the range expected for isolates of the same genotype (≥92%).

In this study, low d_S/d_N values, indicative of positive selection, were observed in certain regions for samples from several patients. Low d_S/d_N values were observed upon analysis of the core gene, where no overlap exists with other HBV genes. The core protein is an important immune target for both antibody and T-cell responses (16), and therefore positive selective

changes within this region are likely an immune evasion strategy of the virus. In general, fewer nonsynonymous than synonymous changes were observed in most regions of the viral genome for samples from all patients. In particular, the reverse transcriptase domains B and C were completely conserved within each matched pair and among all patients, further emphasizing the critical function performed by this region and its requirement for HBV survival (18). Similarly, the X coding region demonstrated relatively negative selection as well as a general lack of nucleotide variation among the matched pairs. The polymerase and X serve essential functions during HBV replication (16), and thus limited nucleotide substitution within these regions would be expected.

The rate of molecular evolution determined by comparing paired samples provides information on the rate of change occurring within a single individual. In particular, the data demonstrate the more rapid or extensive rate of change that occurs in HBeAg-negative patients than in HBeAg-positive patients (11). However, to obtain a more correct mean molecular evolutionary rate for HBV, linear regression analysis of evolutionary distances was performed. This method provides a more accurate value based on a putative ancestral sequence (30), thus avoiding the overestimation resulting from direct comparison of serial sequences (12). Phylogenetic analysis demonstrated that almost all sample pairs clustered uniquely within the tree, indicating an evolutionarily dependent relationship. Serial samples from two study participants (234-6 and 739-11) did not cluster together on the tree, suggesting that the 2004 strain was distinct from the 1979 strain for each pair.

The observation of independent evolution in these two study subjects led us to investigate the contribution of quasiespecies to the HBV molecular evolutionary rate. Immune selective pressure coupled with the lack of proofreading activity by the HBV polymerase likely contributes to the development of quasiespecies complexity and diversity during infection (27). Viral quasiespecies have very closely related genomes but exist in an environment of mutation, selection, and competition, thus creating a dynamic and changing population over time (7). Based on phylogenetic and linear regression analysis of pre-S1/S2 sequence quasiespecies from four sample pairs, it was determined that the mean molecular evolutionary rate did not diverge significantly from the rate calculated using pre-S sequences derived from the directly sequenced or dominant strain. This result suggests that the quasiespecies populations from HBeAg-

negative, asymptomatic chronic HBV carriers during long-term follow-up over 25 years did not contribute significantly to the putative overall evolutionary rate and, therefore, that the dominant sequence is sufficient for rate estimation.

Although the sample pair quasispecies evolutionary rate was not significantly different, the observation that two sample pairs showed independent evolution may be related to quasispecies competition. Replacement of the dominant strain observed in 1979 with a minority quasispecies strain may have occurred due to a selective advantage of the minority variant during the quiescent phase of chronic infection in the study participants (9). For example, the three-codon deletion observed in the pre-S2 region from the majority of 234-6 quasispecies clones from 2004 may contribute a selective advantage to the virus to allow it to become the dominant strain sometime in the future. Such mutations within the pre-S2 region (deletions and start codon mutations) are characteristic of genomes from the HBeAg-negative phase of infection (9). Another explanation for the observed independent evolution may be reinfection with a different HBV genotype B strain during the follow-up period.

The mean nucleotide substitution rate observed in this study was slightly higher than previously estimated rates based on HBeAg-positive carriers, as transmission is assumed to occur predominantly through HBeAg-positive donors (1, 13, 24, 29). However, transmission from carriers negative for HBeAg has been documented (4, 6, 31). The observed evolutionary rate validates previous statements that viruses lacking HBeAg evolve more rapidly, possibly as a function of increased immune pressure during the immune clearance phase of infection (1, 3, 11, 19, 28). The chronic infection phase for all patients investigated in this study was typically quiescent or asymptomatic and HBeAg negative, suggesting a reduction in host immune activity following seroconversion. Therefore, the slightly higher evolutionary rate observed in this study, despite less selection pressure, may be related to the seroconversion event driving quasispecies complexity and diversification. The more diversified quasispecies pool would then undergo competition during the follow-up period to obtain the most "fit," and thus dominant, genome. Furthermore, the reduced selection pressure during the HBeAg-negative chronic phase may allow the accumulation of mutations due to error-prone reverse transcription during replication. Indeed, since overall more synonymous mutations were observed in the coding regions of the study sequences, it is likely that selective outgrowth of sequences having a structure/function advantage for the virus occurred throughout the follow-up period (27).

In conclusion, further analysis of HBV evolutionary patterns should include both HBeAg-positive and -negative symptomatic and asymptomatic patients representing different HBV genotypes to truly characterize HBV sequence divergence over time. In this manner, estimating the molecular clock and origins of HBV may be done more accurately.

REFERENCES

1. Ali Fares, M., and E. C. Holmes. 2002. A revised evolutionary history of hepatitis B virus (HBV). *J. Mol. Evol.* **54**:807–814.
2. Bozkaya, H., U. S. Akarca, B. Ayola, and A. Lok. 1997. High degree of conservation in the hepatitis B virus core gene during the immune tolerant phase in perinatally acquired chronic hepatitis B virus infection. *J. Hepatol.* **26**:508–516.
3. Bozkaya, H., B. Ayola, and A. Lok. 1996. High rate of mutations in the hepatitis B core gene during the immune clearance phase of chronic hepatitis B virus infection. *Hepatology* **24**:32–37.
4. Buster, E., A. A. van der Eijk, and S. W. Schalm. 2003. Doctor to patient transmission of hepatitis B virus: implications of HBV DNA levels and potential new solutions. *Antivir. Res.* **60**:79–85.
5. Cline, J., J. C. Braman, and H. H. Hogrefe. 1996. PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24**:3546–3551.
6. de Arquer, A., J. Pastor, E. Escude, E. Fos, and A. Palomeque. 1985. Vertical transmission of hepatitis B from a maternal HBeAg-negative and anti HBe-positive carrier. *An. Esp. Pediatr.* **22**:33–35.
7. Domingo, E., V. Martín, C. Perales, A. Grande-Pérez, J. García-Arriaza, and A. Arias. 2006. Viruses as quasispecies: biological implications. *Curr. Top. Microbiol. Immunol.* **299**:51–82.
8. Gojobori, T., and S. Yokoyama. 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. *Proc. Natl. Acad. Sci. USA* **82**:4198–4201.
9. Günther, S. 2006. Genetic variation in HBV infection: genotypes and mutants. *J. Clin. Virol.* **36**:S3–S11.
10. Günther, S., B. Li, S. Miska, D. H. Kruger, H. Meisel, and H. Will. 1995. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J. Virol.* **69**:5437–5444.
11. Hannoun, C., P. Horal, and M. Lindh. 2000. Long-term mutation rates in the hepatitis B virus genome. *J. Gen. Virol.* **81**:75–83.
12. Ina, Y., M. Mizokami, K. Ohba, and T. Gojobori. 1994. Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. *J. Mol. Evol.* **38**:50–56.
13. Kidd-Ljunggren, K., Y. Miyakawa, and A. H. Kidd. 2002. Genetic variability in hepatitis B viruses. *J. Gen. Virol.* **83**:1267–1280.
14. Korber, B. 2001. HIV signature and sequence variation analysis, p. 55–72. In A. G. Rodrigo and G. H. Learn (ed.), *Computational analysis of HIV molecular sequences*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
15. Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**:1244–1245.
16. Lee, J. Y., and S. Locarnini. 2004. Hepatitis B virus: pathogenesis, viral intermediates, and viral replication. *Clin. Liver Dis.* **8**:301–320.
17. Lindh, M., J. Gonzalez, G. Norkrans, and P. Horal. 1998. Genotyping of hepatitis B virus by restriction pattern analysis of a pre-S amplicon. *J. Virol. Methods* **72**:163–174.
18. Locarnini, S. 2003. Hepatitis B viral resistance: mechanisms and diagnosis. *J. Hepatol.* **39**:S124–S132.
19. Locarnini, S. 2005. Molecular virology and the development of resistant mutants: implications for therapy. *Semin. Liver Dis.* **25**(Suppl. 1):9–19.
20. Minuk, G. Y., L. E. Nicolle, B. Postl, J. G. Waggoner, and J. H. Hoofnagle. 1982. Hepatitis virus infection in an isolated Canadian Inuit (Eskimo) population. *J. Med. Virol.* **10**:255–264.
21. Mizokami, M., and E. Orito. 1999. Molecular evolution of hepatitis viruses. *Intervirology* **42**:159–165.
22. Mizokami, M., E. Orito, K. Ohba, K. Ikey, J. Y. N. Lau, and T. Gojobori. 1997. Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.* **44**:S83–S90.
23. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
24. Okamoto, H., M. Imai, M. Kametani, T. Nakamura, and M. Mayumi. 1987. Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno-fetal transmission. *Jpn. J. Exp. Med.* **57**:231–236.
25. Osioy, C. 2002. Sensitive detection of HBsAg mutants by a gap ligase chain reaction assay. *J. Clin. Microbiol.* **40**:2566–2571.
26. Rozanov, M., U. Plikat, C. Chappey, A. Kochergin, and T. Tatusova. 2004. A web-based genotyping resource for viral sequences. *Nucleic Acids Res.* **1**:W654–W659.
27. Shuhart, M., D. Sullivan, K. Bekele, R. Harrington, M. Kitahata, T. Mathisen, L. Thomassen, S. Emerson, and D. Gretch. 2006. HIV infection and antiretroviral therapy: effect on hepatitis C virus quasispecies variability. *J. Infect. Dis.* **193**:1211–1218.
28. Simmonds, P. 2001. The origin and evolution of hepatitis viruses in humans. *J. Gen. Virol.* **82**:693–712.
29. Simmonds, P., and S. Midgley. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J. Virol.* **79**:15467–15476.
30. Tanaka, Y., K. Hanada, M. Mizokami, A. Yeo, J. Shih, T. Gojobori, and H. Alter. 2002. A comparison of the molecular clock of hepatitis C virus in the

- United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. *Proc. Natl. Acad. Sci. USA* **99**:15584–15589.
31. **The Incident Investigation Teams and Others.** 1997. Transmission of hepatitis B to patients from four infected surgeons without hepatitis B e antigen. *N. Engl. J. Med.* **336**:178–184.
32. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**:4876–4882.
33. **Whalley, S. A., D. Brown, G. J. M. Webster, R. Jacobs, S. Reignat, A. Bertolotti, C. Teo, V. Emery, and G. M. Dusheiko.** 2004. Evolution of hepatitis B virus during primary infection in humans: transient generation of cytotoxic T-cell mutants. *Gastroenterology* **127**:1131–1138.